

SiteScan User Guide

Table of contents

Welcome	3
Overview	3
Table of Contents.....	5
User Interface	6
Tool Ribbon	7
Results Grid	10
Properties Panel	14
Out-Links Panel.....	15
In-Links Panel	18
Images Panel	20
Styles Panel	22
Scripts Panel	24
Emails Panel	26
Source Panel.....	27
Options.....	29
Scanning	30
Export	34
Licensing	36

Welcome

Welcome

[Overview](#)

[Table of Contents](#)

Overview

Introduction

SiteScan is a Windows desktop program which checks websites for broken links and analyzes page performance.

From a starting URL, SiteScan will spider and analyze every page or file on a web site or intranet. It will find issues such as slow, mistyped or broken links and allow the user to drill down to perform detailed page analysis and locate the relevant code in the HTML source. Results are displayed in real-time and can be interacted with immediately as scanning continues in the background.

Intuitive and ready to run out-of-the box, SiteScan is easy to use and highly configurable. It provides many user defined options such as scan depth, timeouts, scheme/protocol and file type exclusions rules. It runs from any Windows desktop PC or server and communicates directly with the website being scanned. SiteScan does not use any intermediate cloud based services during processing for complete data privacy.

Available in both free and registered editions, SiteScan Broken Link Checker is licensed to an individual user with heavily discounted bulk licensing available for corporate and educational users. The free edition of SiteScan is provided for non-business use.



Key Features

- Spider websites for slow or broken links.
- Find mistyped or badly formatted URLs
- Check any type of web page, file or media.
- Spider all hyperlinks or only first occurrence.
- Control scan depth, page & link count limits.
- Scan http, https, file, mailto & javascript schemes.
- Server & client-side redirect support.
- Spider intranet file:// linked content.
- Measure response, download and HTML load time.
- Set spider speed, pause and resume.
- Review page titles, description and alt text.
- Set timeouts to highlight slow loading pages.
- Save, load & refresh existing results.
- View all in-links and out-links for any page or
- Highlight results with conditional formatting.
- Work with results in real-time during scanning.
- Sort, group, search & filter results like Excel.
- One-click filtering for all errors & warnings.
- Highlight URL and image markup in HTML source.
- Export spider results to XLSX, XLS, CSV.

For more information visit: www.byteshift.co.uk or email support@byteshift.co.uk

Table of Contents

Welcome

[Overview](#)
[Table of Contents](#)

User Interface

[Tool Ribbon](#)
[Results Grid](#)
[Properties Panel](#)
[Out-Links Panel](#)
[In-Links Panel](#)
[Images Panel](#)
[Styles Panel](#)
[Scripts Panel](#)
[Emails Panel](#)
[Source Panel](#)

Options

[Scanning](#)
[Export](#)
[Licensing](#)

User Interface

User Interface

[Tool Ribbon](#)

[Results Grid](#)

[Properties Panel](#)

[Out-Links Panel](#)

[In-Links Panel](#)

[Images Panel](#)

[Styles Panel](#)

[Scripts Panel](#)

[Emails Panel](#)

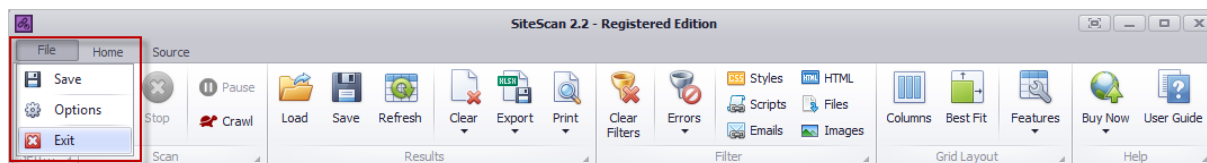
[Source Panel](#)

Tool Ribbon

Quick Reference

The following quick reference provides an overview for each menu and icon on the SiteScan tool ribbon. Hover over any menu item within the application to see this information in a tool-tip.

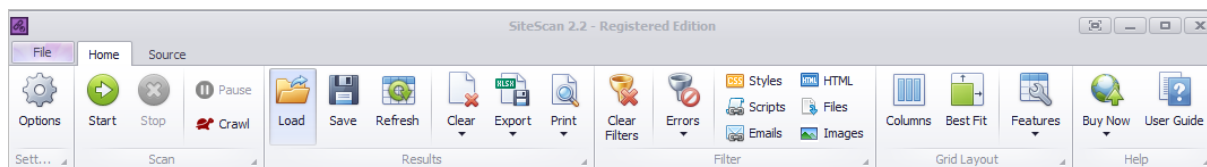
File Menu



File
Save Save current results to file.
Options Display the Options dialog.
Exit Exit SiteScan (prompts to Save if required)

Home Ribbon

The Home ribbon provides access to the most frequently used features, including: starting and stopping scans, filtering data and manipulating results in the grid.

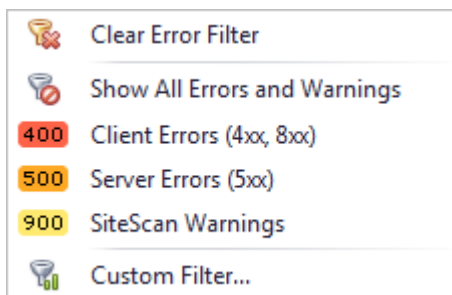


Settings	Scan	Results	Filter	Grid Layout	Help
Options Display the Options dialog.	Start Start a new scan.	Load Load existing scan results from file.	Clear Filters Removes any applied filters to display all results.	Column Choose link information columns to display on the results grid.	Buy Now Visit the SiteScan website to purchase a license key or enter a purchased key to activate SiteScan.
	Stop Stop a running scan.	Save Save current results to file.	Errors Allows results to be filtered for all or only specific categories of errors. See the Error Filters section below.	Best Fit Auto-size column widths to fit screen.	User Guide Display the SiteScan user guide.
	Pause Pause / resume scanning.	Refresh Refresh performance and status for existing results without spidering.	Styles Filters the results to display on linked CSS stylesheets.	Features Show / hide grid features such as Group By, Find Panel or Filter Rows.	
	Crawl Slow down a running scan.	Clear Clear all existing results or only performance and status data.	Scripts Filters the results to displayed only linked script files (e.g. JavaScript)		
		Export Export results to Excel or CSV format.	Emails Filters the results to display only email addresses specified		

			using the mailto: scheme.		
		Print Print or print-preview results.	HTML Filters the results to display HTML pages. Filtering is based on the content type header returned from the server.		
			Files Filters the results to display linked files by looking in the content type header for keys such as application, binary and audio.		
			Images Filters the results to display only images linked to by style markup.		

Error Filters

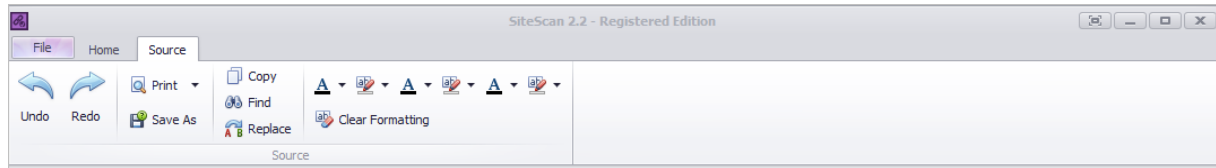
The **Error Filter** menu provides predefined filters to allow quick viewing of results with specific HTTP Status Codes or warning messages.






Error Filter
Clear Error Filter Clear any applied filters and displays all results.
Show All Errors and Warnings Filters the results to show only those with Client Errors, Server Errors and SiteScan warnings.
Client Errors Filters the results to show only Client Errors beginning with the 400 or 800 HTTP status code.
Server Errors Filters the results to show only Server Errors beginning with the 500 HTTP status code.
SiteScan Warnings Filters the results to show only warning messages suggested by SiteScan such as slow loading pages or possible mistyped URLs.
Custom Filter Displays the filter creation pop-up window to allow creation of user defined, complex filter expressions on multiple columns.

Source Ribbon

The Source ribbon work exclusively with the Source Code preview panel and provides access to edit, format, save and print source code.



Source	
Undo	Undo the last text or formatting change to the source code.
Redo	Redo the last text or formatting change to the source code.
Print	Print or Print Preview the source code with formatting.
Save As	Save the source code to file.
Copy	Copy any selected text to the Windows clipboard.
Find	Find and highlight text in the source code.
Replace	Find and replace text in the source code.
	Change selected text font color.
	Change selected text highlight color.
	Clear formatting from selected text.

Results Grid

Results Grid

The Results Grid is used to display and interact with scan results. Results are displayed in real-time as a scan runs or existing results loaded from file for further analysis or refresh.

Overview

Results are displayed in an Excel-like grid with full sort, filtering and grouping capabilities. The most common features are described below with many additional features accessible by right-clicking on various places within the grid, such as column headers, cells or group headers.

Drag a column header here to group by that column

Source Page		Target Link							
ID	Source Page URL	Link Text	Link Relative URL	Resp... Code	Response Feedback	Response Time (s)	Content Type	Size (kb)	
1			/?p=us	OK	OK	0.30	text/html; ...	0	
2	https://uk.yahoo.com/?p=us	Việt Nam	https://vn.yahoo.com/?p=dnr	OK	OK	0.86	text/html; ...	0	
3	https://uk.yahoo.com/?p=us	Malaysia	https://malaysia.yahoo.com/?p=dnr	OK	OK	0.78	text/html; ...	0	
4	https://uk.yahoo.com/?p=us	Ireland	https://ie.yahoo.com/?p=dnr	OK	OK	0.83	text/html; ...	0	
5	https://uk.yahoo.com/?p=us	Indonesia	https://id.yahoo.com/?p=dnr	OK	OK	0.80	text/html; ...	0	
6	https://uk.yahoo.com/?p=us	Belgique (fr)	https://fr-be.yahoo.com/?p=dnr	OK	OK	0.87	text/html; ...	0	
7	https://uk.yahoo.com/?p=us	Ελλάδα	https://gr.yahoo.com/?p=dnr	OK	OK	0.88	text/html; ...	0	
8	https://uk.yahoo.com/?p=us	India	https://in.yahoo.com/?p=dnr	OK	OK	0.81	text/html; ...	0	
9	https://uk.yahoo.com/?p=us	Philippines	https://ph.yahoo.com/?p=dnr	OK	OK	0.91	text/html; ...	0	
10	https://uk.yahoo.com/?p=us	België (nl)	https://be.yahoo.com/?p=dnr	OK	OK	0.87	text/html; ...	0	
11	https://uk.yahoo.com/?p=us	Deutschland	https://de.yahoo.com/?p=dnr	OK	OK	0.87	text/html; ...	0	

Record 1 of 676

Hover over any column heading for a detailed description of the information the columns contains:

Target Link				
Response Code	Response Feedback		Response Time (s)	Content Type
OK	The HTTP response code returned from the web server when the link was requested		0.65	text/html; ...
OK	OK		0.28	text/html; ...
OK	OK		0.32	text/html; ...
OK	OK		0.30	text/html; ...

Column Groups (important)

All columns which can be displayed in the results grid belong to one of two groups (also referred to as bands). These are:

- Source Page
- or
- Target Link

Source Page		Target Link		
ID	Source Page URL	Link Text	Link Relative URL	Response Code

Columns which appear in the Source Page group contain information and links pertaining to the source or referring page on which the Target Links were found. Clicking a URL in the Source Page

group will load the referring page in to the preview Browser then load its properties links, images, etc, for examination.

Columns which appear in the Target Links group contain information and the links found ON the source or referring page. Clicking a URL in the Target Links group will load the page that link points to (the target) in to the preview Browser then load it's properties links, images, etc for examination.

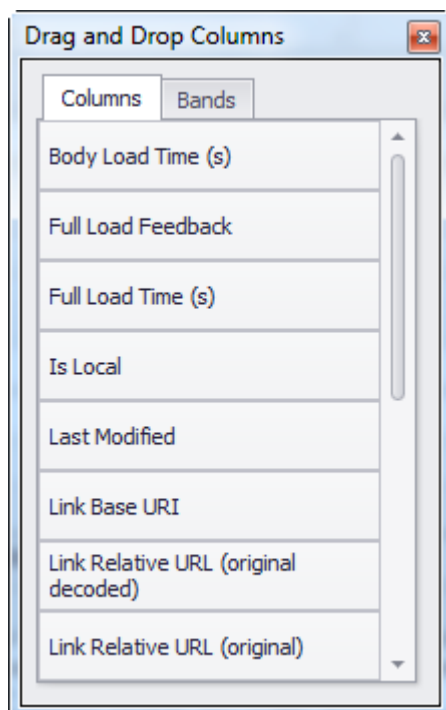
This distinction is important. For example, when you wish to find and highlight a broken link in Browser or Source Code, you must load the **Source Page**, as this is the page on which the broken link can be found, then choose the link to highlight from the **Links** panel. For more information on highlighting links, see [Links Panel](#) or [Browser Preview](#).

Choosing Columns

Several columns are available to be shown on the grid, each displaying different information about the page or file which was scanned. Several of the most frequently used columns are shown by default, with other columns added via the **Columns** button in the **Grid Layout** section of the toolbar ribbon.

To add columns:

1. Click the **Columns** button on the toolbar ribbon to display the list of available columns.
2. Drag a column from the pop-up window to the grid



To remove columns:

1. Click and drag a column heading off the grid. The cursor will change to a large X when the column can be dropped to be removed.

Previewing a Result

Individual results can be inspected in detail using the Page Analysis tabs below the main results grid.

To preview a result, select it by clicking its row in the results grid.

Filtering

Results can be filtered in several ways:

- By selecting a predefined error filter from the **Filter Errors** drop-down button in the **Filters** section of the tool ribbon.
- By selecting a predefined content type filter from the **Filters** section of the toolbar ribbon (such as **Email**, **Styles** or **Scripts**).
- By typing in to the auto-filter row immediately below the column header.
- By hovering over a column header to display the drop-down filter icon then selecting a column value to filter on.

Sorting

Results are sorted by clicking on a column header. Clicking the header multiple times will toggling sorting between ascending, descending and unsorted.

Grouping

Results are grouped by dragging a column header to the **Group By** row shown immediately above the column headers. Nested groupings can be created by dragging multiple column headers to the **Group By** row. To remove a grouping, drag the column header out of the **Group By** row.

Saving

Results are saved by clicking the **Save** icon on the **Results** section of the tool ribbon. Results are save in XML format.

Two files are created:

- A file with a .xml extension contains all scan results (filtering is ignored when saving)
- A file with a .xmlr extensions contains all referrer data (all pages found which point to links in the results)

When backing up, copying or moving results both files should be moved and kept together. Note: The main .xml results can be loaded without referrer data, but any referrers will not be displayed in the In-Links panel.

Loading

Results are loaded by clicking the **Load** icon on the **Results** section of the tool ribbon.

When loading SiteScan will first load the main .xml results file, then any referrer data from the .xmlr file if found.

Exporting

Results can be exported for external analysis in three formats:

- Excel .xlsx format
- Excel .xls format

- Comma separated .csv format

The .csv format is fastest as cell formatting is removed.

Properties Panel

Properties Panel

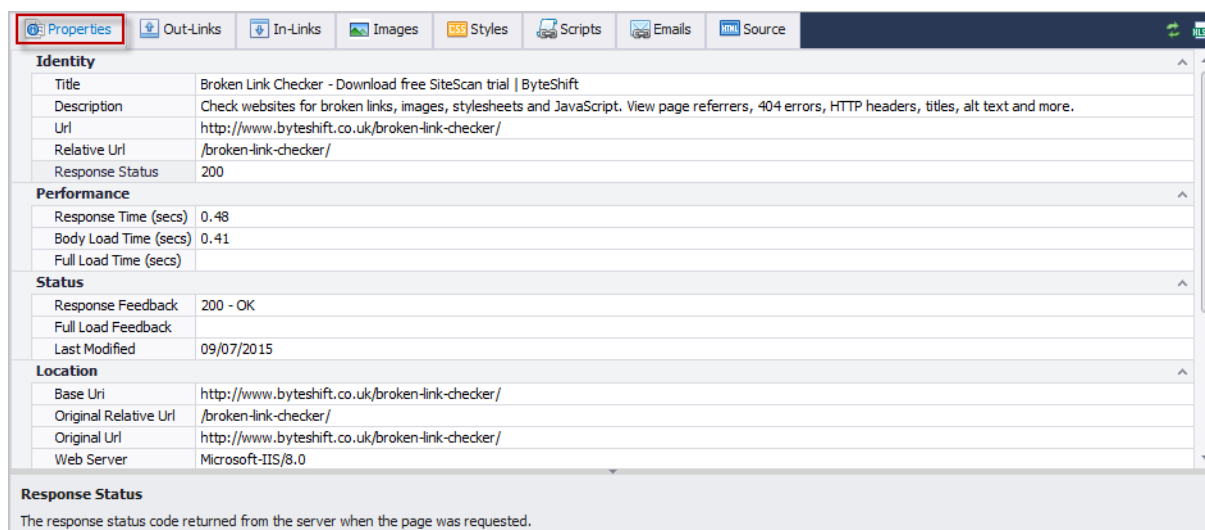
The Properties panel displays meta-data and performance measurements for the page currently selected in the main crawl Results Grid. Properties are refreshed, loaded and displayed whenever a page is selected in the results grid.

Overview

The properties displayed in Properties panel and the columns shown on the Results Grid are the same. When a row is selected in the Results Grid the data obtained during spidering is displayed for the selected result.

Properties are grouped in to categories with each displaying related information. Values are read-only, but can be selected and copied to the Windows clipboard.

By default the properties for each link is only refreshed whenever the page is selected in the main Result Grid. See the Refreshing section below for information on refreshing properties.



Properties Panel	
Identity <ul style="list-style-type: none"> Title: Broken Link Checker - Download free SiteScan trial ByteShift Description: Check websites for broken links, images, stylesheets and JavaScript. View page referrers, 404 errors, HTTP headers, titles, alt text and more. Url: http://www.byteshift.co.uk/broken-link-checker/ Relative Url: /broken-link-checker/ Response Status: 200 	
Performance <ul style="list-style-type: none"> Response Time (secs): 0.48 Body Load Time (secs): 0.41 Full Load Time (secs): 	
Status <ul style="list-style-type: none"> Response Feedback: 200 - OK Full Load Feedback: Last Modified: 09/07/2015 	
Location <ul style="list-style-type: none"> Base Uri: http://www.byteshift.co.uk/broken-link-checker/ Original Relative Url: /broken-link-checker/ Original Url: http://www.byteshift.co.uk/broken-link-checker/ Web Server: Microsoft-IIS/8.0 	
Response Status The response status code returned from the server when the page was requested.	

Refreshing

Properties can be refresh directly from the web server by either:

- Double-clicking the related row in the main Results Grid.
- Clicking the green Refresh icon to the right of the tab headings.

Exporting

Properties can be exported to Excel in XLSX format.

- Click the XLSX icon to the right of the tab headings.

Out-Links Panel

Out-Links Panel

The Out-Links panel displays all internal or external hyperlinks to other HTML pages (or files) from the page currently selected in the main crawl Results Grid. Out-Links are loaded and displayed whenever a page is selected in the results grid.

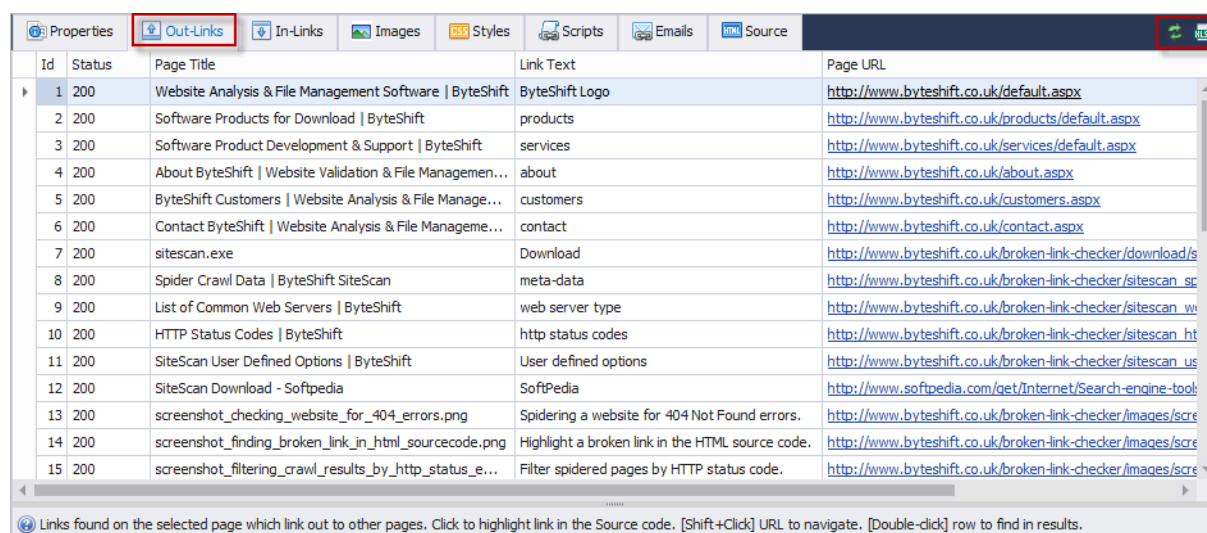
Overview

The links displayed in the Out-Links panel are obtained at the point the page was crawled and take in to account any spider rules specified in options (such as scheme or file type filtering specified). Therefore, the Out-Links listed may only be a subset of all links on the web page. Out-Links are only listed once in this view regardless of how many times they may appear on a page.

Each Out-Link displays the HTML Title of the page to which the link points, the Link Text used in the anchor tag and the absolute URL of the link. The Relative URL is also available to be shown by right-clicking on the grid column headings.

Out-Links are cross referenced with available scan results to show their current HTTP status (e.g. OK, 404 Not Found, 500, etc) and are color coded to highlight errors and warnings. As a scan runs and more scan results become available, the appropriate status will be shown against each link.

By default the HTTP status for each Out-Link is only refreshed whenever the page is selected in the main Result Grid. See the Refreshing section below for information on refreshing status.



Id	Status	Page Title	Link Text	Page URL
1	200	Website Analysis & File Management Software ByteShift	ByteShift Logo	http://www.byteshift.co.uk/default.aspx
2	200	Software Products for Download ByteShift	products	http://www.byteshift.co.uk/products/default.aspx
3	200	Software Product Development & Support ByteShift	services	http://www.byteshift.co.uk/services/default.aspx
4	200	About ByteShift Website Validation & File Managemen...	about	http://www.byteshift.co.uk/about.aspx
5	200	ByteShift Customers Website Analysis & File Manage...	customers	http://www.byteshift.co.uk/customers.aspx
6	200	Contact ByteShift Website Analysis & File Manageme...	contact	http://www.byteshift.co.uk/contact.aspx
7	200	sitescan.exe	Download	http://www.byteshift.co.uk/broken-link-checker/download/s
8	200	Spider Crawl Data ByteShift SiteScan	meta-data	http://www.byteshift.co.uk/broken-link-checker/sitescan_sc
9	200	List of Common Web Servers ByteShift	web server type	http://www.byteshift.co.uk/broken-link-checker/sitescan_w
10	200	HTTP Status Codes ByteShift	http status codes	http://www.byteshift.co.uk/broken-link-checker/sitescan_ht
11	200	SiteScan User Defined Options ByteShift	User defined options	http://www.byteshift.co.uk/broken-link-checker/sitescan_us
12	200	SiteScan Download - Softpedia	SoftPedia	http://www.softpedia.com/get/Internet/Search-engine-tool
13	200	screenshot_checking_website_for_404_errors.png	Spidering a website for 404 Not Found errors.	http://www.byteshift.co.uk/broken-link-checker/images/scr
14	200	screenshot_finding_broken_link_in_html_sourcecode.png	Highlight a broken link in the HTML source code.	http://www.byteshift.co.uk/broken-link-checker/images/scr
15	200	screenshot_filtering_crawl_results_by_http_status_e...	Filter spidered pages by HTTP status code.	http://www.byteshift.co.uk/broken-link-checker/images/scr

Links found on the selected page which link out to other pages. Click to highlight link in the Source code. [Shift+Click] URL to navigate. [Double-click] row to find in results.

Navigating to an Out-Link in Results

To jump to the corresponding result in the main Results Grid for any link:

- Double-click the link's row.

Previewing an Out-Link Result in the Browser

To open a link in the browser:

- Hold the **[Shift]** key and click the URL in the **Page URL** column.

Highlighting Out-Links in the HTML Source

Links can be highlighted in the **Source** code preview tab:

- Click any link row to select it.
- Click the **Source** tab to view the source.


<div> <div>Properties</div> <div>Out-Links</div> <div>In-Links</div> <div>Images</div> <div>CSS Styles</div> <div>Scripts</div> <div>Emails</div> <div>HTML Source</div> </div>					
Id		Status	Page Title	Link Text	Page URL
1	200		Website Analysis & File Management Software ByteShift	ByteShift Logo	http://www.byteshift.co.uk/default.aspx
2	200		Software Products for Download ByteShift	products	http://www.byteshift.co.uk/products/default.aspx
3	200		Software Product Development & Support ByteShift	services	http://www.byteshift.co.uk/services/default.aspx
4	200		About ByteShift Website Validation & File Managemen...	about	http://www.byteshift.co.uk/about.aspx
5	200		ByteShift Customers Website Analysis & File Manage...	customers	http://www.byteshift.co.uk/customers.aspx
6	200		Contact ByteShift Website Analysis & File Manageme...	contact	http://www.byteshift.co.uk/contact.aspx
7	200		sitescan.exe	Download	http://www.byteshift.co.uk/broken-link-checker/download/s
8	200		Spider Crawl Data ByteShift SiteScan	meta-data	http://www.byteshift.co.uk/broken-link-checker/sitescan_sp
9	200		List of Common Web Servers ByteShift	web server type	http://www.byteshift.co.uk/broken-link-checker/sitescan_w
10	200		HTTP Status Codes ByteShift	http status codes	http://www.byteshift.co.uk/broken-link-checker/sitescan_ht
11	200		SiteScan User Defined Options ByteShift	User defined options	http://www.byteshift.co.uk/broken-link-checker/sitescan_us
12	200		SiteScan Download - Softpedia	SoftPedia	http://www.softpedia.com/get/Internet/Search-engine-tool
13	200		screenshot_checking_website_for_404_errors.png	Spidering a website for 404 Not Found errors.	http://www.byteshift.co.uk/broken-link-checker/images/scre
14	200		screenshot_finding_broken_link_in_html_sourcecode.png	Highlight a broken link in the HTML source code.	http://www.byteshift.co.uk/broken-link-checker/images/scre
15	200		screenshot_filtering_crawl_results_by_http_status_e...	Filter spidered pages by HTTP status code.	http://www.byteshift.co.uk/broken-link-checker/images/scre

A number will be appended to the **Source** tab caption to indicate how many occurrences of the selected link were found. Each occurrence of the link found is highlighted in yellow.

```
107      <tr>
108          <td>
109              <table class="headerContent " >
110                  <tr>
111                      <td>
112                          <div class="headerLogo"><a href="/default.aspx" title="ByteShift"></a></div>
114                      <div class="headerNavigation">
115                          <div class="linkMargin"><a id="elmLinkHome" class="headerLink"
116                          href="/default.aspx">home</a></div>
117                          <span class="headerLinkSeparator">|</span>
118                          <div class="linkMargin"><a id="elmLinkProducts"
119                          class="headerLinkSelected" href="/products/default.aspx">products</a></div>
120                          <span class="headerLinkSeparator">|</span>
121                          <div class="linkMargin"><a id="elmLinkServices" class="headerLink"
122                          href="/services/default.aspx">services</a></div>
123                          <span class="headerLinkSeparator">|</span>
124                          <div class="linkMargin"><a id="elmLinkAbout" class="headerLink"
125                          href="/about.aspx">about</a></div>
126                          <span class="headerLinkSeparator">|</span>
127                          <div class="linkMargin"><a id="elmLinkCustomers" class="headerLink"
128                          href="/customers.aspx">customers</a></div>
129                          <span class="headerLinkSeparator">|</span>
```

Refreshing

Out-Links can be refresh for the current page directly from the web server by either:

- Double-clicking the related row in the main Results Grid.
- Clicking the green Refresh icon  to the right of the tab headings.

Exporting

Out-Links can be exported to Excel in XLSX format.

- Click the XLSX icon  to the right of the tab headings.

In-Links Panel

In-Links Panel

The In-Links panel lists all pages with links pointing to the page or file currently selected in the main Results Grid.

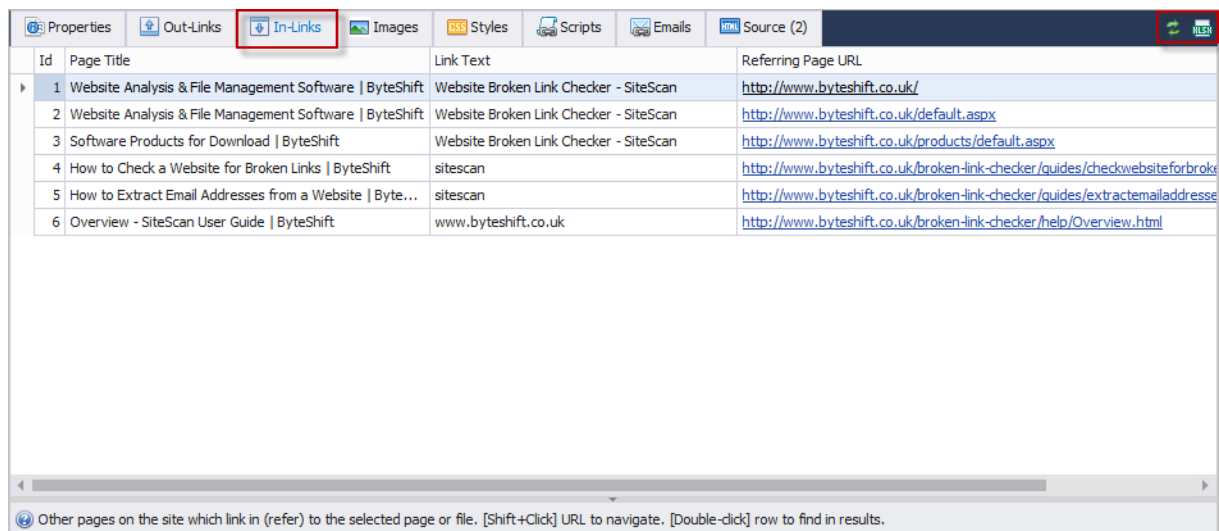
Overview

The list of In-Links displays the Page Title, Link Text and absolute URL of all source pages found during scanning which contain a link to the selected page or file. In-Links are subject to any scan depth and filtering rules specified in [Options](#).

For example:

If a scan is performed on the URL `www.mysite.com/products` with the scan depth filter set to only extract and follow links "On the same level of beneath the starting URL", then only referring pages at or below the URL `www.mysite.com/products` will be reported.

By default the In-Links for each page or file are only refreshed whenever the page is selected in the main Result Grid. See the Refreshing section below for information on refreshing In-Link data.



Id	Page Title	Link Text	Referring Page URL
1	Website Analysis & File Management Software ByteShift	Website Broken Link Checker - SiteScan	http://www.byteshift.co.uk/
2	Website Analysis & File Management Software ByteShift	Website Broken Link Checker - SiteScan	http://www.byteshift.co.uk/default.aspx
3	Software Products for Download ByteShift	Website Broken Link Checker - SiteScan	http://www.byteshift.co.uk/products/default.aspx
4	How to Check a Website for Broken Links ByteShift	sitescan	http://www.byteshift.co.uk/broken-link-checker/guides/checkwebsiteforbrok
5	How to Extract Email Addresses from a Website Byte...	sitescan	http://www.byteshift.co.uk/broken-link-checker/guides/extractemailaddresse
6	Overview - SiteScan User Guide ByteShift	www.byteshift.co.uk	http://www.byteshift.co.uk/broken-link-checker/help/Overview.html

Other pages on the site which link in (refer) to the selected page or file. [Shift+Click] URL to navigate. [Double-click] row to find in results.

Navigating to a Referrer in Results

To jump to the corresponding result in the main Results Grid for any referrer:

- Double-click the referrer row.


Previewing a Referrer in the Browser

To open the referring page in the browser:

- Hold the **[Shift]** key and click the URL in the **Referring Page URL** column.

Refreshing

The list of In-Links can be refresh from the current crawl results by either:

- Double-clicking the related row in the main Results Grid.
- Clicking the green Refresh icon  to the right of the tab headings.

Exporting

In-Links can be exported to Excel in XLSX format.

- Click the XLSX icon  to the right of the tab headings.

Images Panel

Images Panel

The Images panel lists images declared using `` on the page currently selected in the main Results Grid. Images are displayed whenever a page is selected in the results grid.

Overview

Images are only listed once in this view regardless of how many times they may appear on a page. The list of images displays the ALT text for the first occurrence of the image found on the page along with the Absolute URL to the image file. The Relative URL is also available to be shown by right-clicking on the grid column headings.

<div> <div>Properties</div> <div>Out-Links</div> <div>In-Links</div> <div>Images</div> <div>Styles</div> <div>Scripts</div> <div>Emails</div> <div>Source (2)</div> </div>				
Id		Status	Alt Text	Image URL
▶	1	200	ByteShift Logo	http://www.byteshift.co.uk/images/logos/byteshift_205x41.png
	2	200	SiteScan Logo	http://www.byteshift.co.uk/images/logos/sitescan_105x105.png
	3	200	Spidering a website for 404 Not Found errors.	http://www.byteshift.co.uk/broken-link-checker/images/screenshots/thumbnail_checking_website_f...
	4	200	Highlight a broken link in the HTML source code.	http://www.byteshift.co.uk/broken-link-checker/images/screenshots/thumbnail_finding_broken_link...
	5	200	Filter spidered pages by HTTP status code.	http://www.byteshift.co.uk/broken-link-checker/images/screenshots/thumbnail_filtering_crawl_resul...
	6	200	Viewing all referring pages (incoming links) pointing to a...	http://www.byteshift.co.uk/broken-link-checker/images/screenshots/thumbnail_finding_all_incoming...
	7	200	Grouping spidered web pages by HTTP status code to f...	http://www.byteshift.co.uk/broken-link-checker/images/screenshots/thumbnail_grouping_spidered ...
	8	200	Grouping crawled web pages by the HTTP content type...	http://www.byteshift.co.uk/broken-link-checker/images/screenshots/thumbnail_grouping_spidered ...
	9	200	Filtering spider results to display email addresses.	http://www.byteshift.co.uk/broken-link-checker/images/screenshots/thumbnail_filtering_crawled pa...
	10	200	Exporting site crawl results to Microsoft Excel.	http://www.byteshift.co.uk/broken-link-checker/images/screenshots/thumbnail_exporting_broken li...
	11	200	Download SiteScan from www.byteshift.co.uk	http://www.byteshift.co.uk/images/byteshift_direct_download.png
	12	200	Comodo Secure Logo	http://www.byteshift.co.uk/images/comodo_secure_90x59.png
	13	200	Yes	http://www.byteshift.co.uk/images/validation_check.png
	14	200	USD Pricing	http://www.byteshift.co.uk/images/flag_usa.png
	15	200	GBP Pricing	http://www.byteshift.co.uk/images/flag_gb.png
	16	200	Buy Now	http://www.byteshift.co.uk/images/buy_now_106x38.png

Images found on the selected page. Click to highlight image in the Source code. [Shift+Click] URL to navigate. [Double-click] row to find in results.

Previewing an Image in the Browser

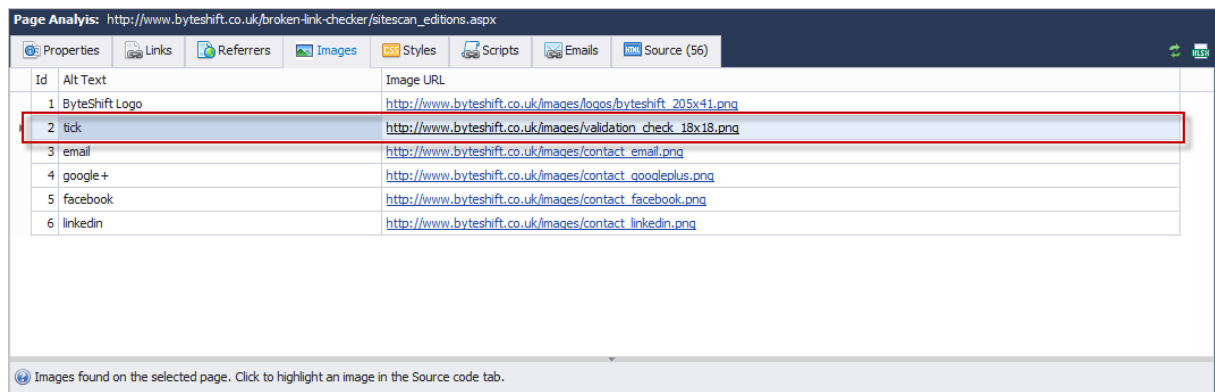
To open an image in the browser (or native associated application):

- Hold the **[Shift]** key and click the URL in the **Image URL** column.

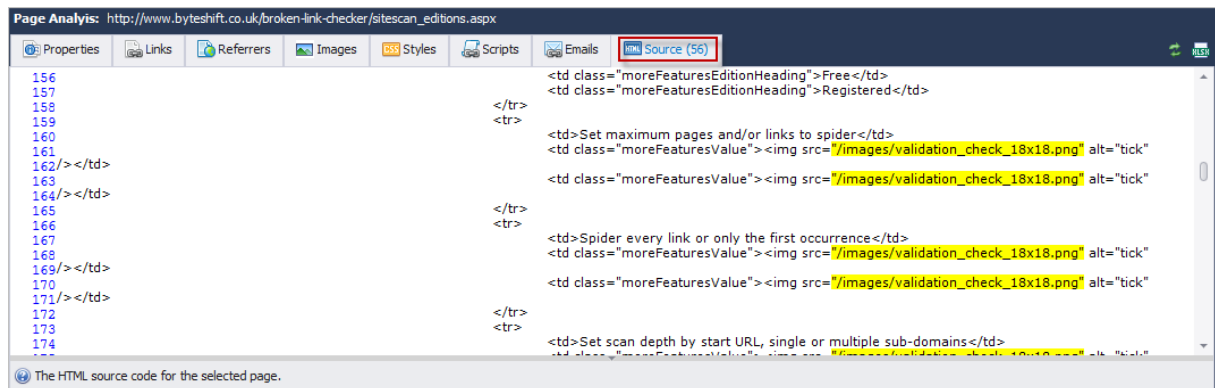
Highlighting Images in the HTML Source

Images can be highlighted in the **Source** code preview tab:

- Click any image row to select it.
- Click the Source tab to view the source.




A number will be appended to the **Source** tab caption to indicate how many occurrences of the selected image were found. Each image found is highlighted in yellow.



Refreshing

The list of images found on the page can be refresh directly from the web server by either:

- Double-clicking the related row in the main Results Grid.
- Clicking the green Refresh icon  to the right of the tab headings.

Exporting

The list of images can be exported to Excel in XLSX format.

- Click the XLSX icon  to the right of the tab headings.

Styles Panel

Styles Panel

The Styles panel lists all style-sheet type links found on the page currently selected in the main Results Grid.

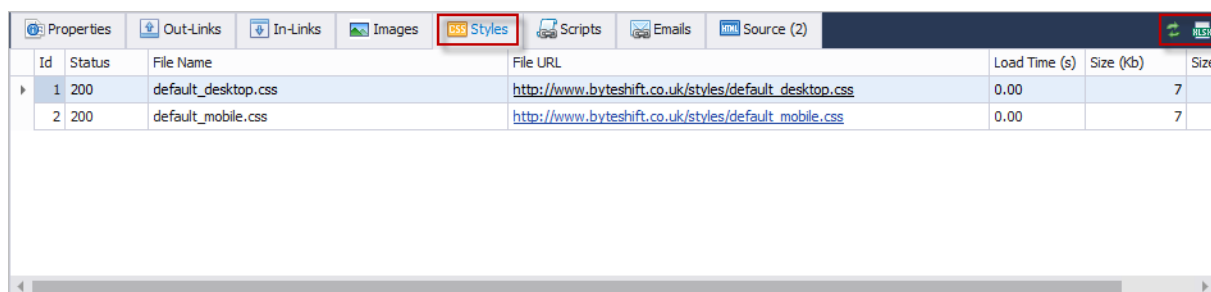
Overview

The list of style-sheets displayed are obtained at the point the page was crawled and take in to account any spider rules specified in options (such as scheme or file type filtering specified). Therefore, those listed may only be a subset of all style-sheets on the web page. Each style-sheets is only listed once in this view regardless of how many times it may appear in a page.

Each style-sheet link displays its HTTP status, file name, absolute URL, load time in seconds and size in both bytes and kilobytes.

Each style-sheet link is cross referenced with the available scan results to show its current HTTP status (e.g. OK, 404 Not Found, 500, etc) and are color coded to highlight errors and warnings. As a scan runs and more scan results become available, the appropriate status will be shown against each link when refreshed.

By default the HTTP status for each style-sheet is only refreshed whenever its page is selected in the main Result Grid. See the Refreshing section below for information on refreshing style-sheet status.



Id	Status	File Name	File URL	Load Time (s)	Size (Kb)	Size
1	200	default_desktop.css	http://www.byteshift.co.uk/styles/default_desktop.css	0.00	7	7
2	200	default_mobile.css	http://www.byteshift.co.uk/styles/default_mobile.css	0.00	7	7

Navigating to a StyleSheet in Results

To jump to the corresponding result in the main Results Grid for any style-sheet:

- Double-click the style-sheet row.

Previewing a StyleSheet in the Browser

To open the style-sheet page in the browser:

- Hold the **[Shift]** key and click the URL in the **File URL** column.

Refreshing

The list of style-sheets can be refresh from the page on the web server by either:

- Double-clicking the related page row in the main Results Grid.
- Clicking the green Refresh icon  to the right of the tab headings.

Exporting

The list of style-sheets can be exported to Excel in XLSX format.

- Click the XLSX icon  to the right of the tab headings.

Scripts Panel

Scripts Panel

The Scripts panel lists all JavaScript type links found on the page currently selected in the main Results Grid.

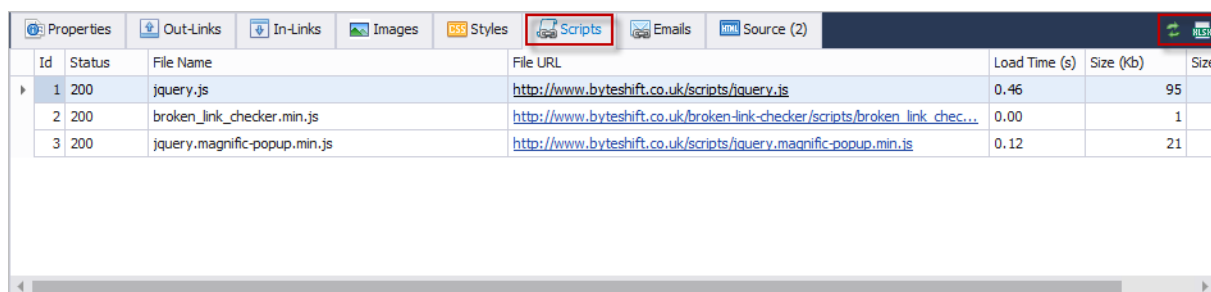
Overview

The list of script files displayed is obtained at the point the page was crawled and takes in to account any spider rules specified in options (such as scheme or file type filtering specified). Therefore, those listed may only be a subset of all scripts on the web page. Each script is only listed once in this view regardless of how many times it may appear in a page.

Each script link displays its HTTP status, file name, absolute URL, load time in seconds and size in both bytes and kilobytes.

Each script link is cross referenced with the available scan results to show its current HTTP status (e.g. OK, 404 Not Found, 500, etc) and color coded to highlight errors and warnings. As a scan runs and more scan results become available, the appropriate status will be shown against each file when refreshed.

By default the HTTP status for each script link is only refreshed whenever its page is selected in the main Result Grid. See the Refreshing section below for information on refreshing script link status.



Id	Status	File Name	File URL	Load Time (s)	Size (Kb)	Size
1	200	jquery.js	http://www.byteshift.co.uk/scripts/jquery.js	0.46	95	
2	200	broken_link_checker.min.js	http://www.byteshift.co.uk/broken-link-checker/scripts/broken_link_chec...	0.00	1	
3	200	jquery.magnific-popup.min.js	http://www.byteshift.co.uk/scripts/jquery.magnific-popup.min.js	0.12	21	

Navigating to a Script File in Results

To jump to the corresponding result in the main Results Grid for any script link:

- Double-click the script link row.

Previewing a Script File in the Browser

To open the script file in the browser (or other native associated application):

- Hold the **[Shift]** key and click the URL in the **File URL** column.

Refreshing

The list of script links can be refresh from the page on the web server by either:

- Double-clicking the related page row in the main Results Grid.
- Clicking the green Refresh icon  to the right of the tab headings.

Exporting

The list of scripts can be exported to Excel in XLSX format.

- Click the XLSX icon  to the right of the tab headings.

Emails Panel

Emails Panel

The Emails panel lists all email addresses, linked using the mailto: scheme prefix, found on the page currently selected in the main Results Grid.

Overview

The list of email addresses is obtained at the point the page was crawled. Each email address is only listed once in this view regardless of how many times it may appear in a page.

Each email link displays a status (OK or a SiteScan 9xx warning where it appears to be mistyped), link text and the email link as specified in the href attribute of the anchor tag in which it was found.

Properties

Out-Links

In-Links

Images

Styles

Scripts

Emails

Source

Id		Status	Link Text	Email Link
1	OK	inspired@byteshift.co.uk	mailto:inspired@byteshift.co.uk	
2	OK	email	mailto:support@byteshift.co.uk	

Navigating to a Email Address in Results

To jump to the corresponding result in the main Results Grid for any email address:

- Double-click the email link row.

Refreshing

The list of email addresses can be refresh from the page on the web server by either:

- Double-clicking the related page row in the main Results Grid.
- Clicking the green Refresh icon  to the right of the tab headings.

Exporting

The list of emails addresses can be exported to Excel in XLSX format.

- Click the XLSX icon  to the right of the tab headings.

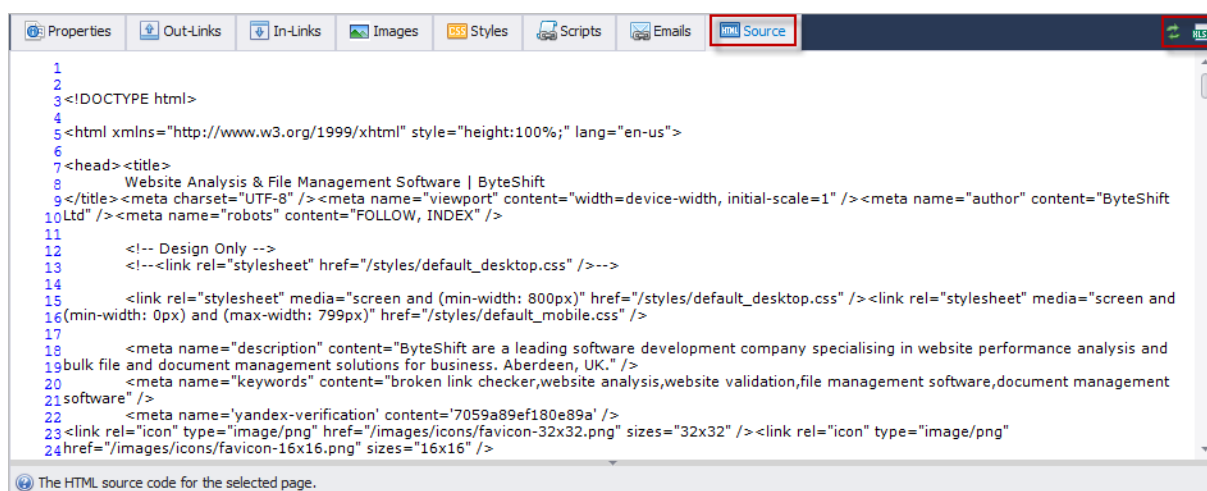
Source Panel

Source Preview

The Source preview is used to display the source code from the page currently selected for analysis in the main Results Grid. It provides highlighting of links as well as editing, formatting and printing of source code. The [Source](#) toolbar ribbon provides features for working with source code.

Overview

The Source preview is a rich text editor display with support for line numbering, text highlighting, formatting and saving. Its contents are synchronized with the page selected in the Results Grid as well as Properties, Out-Links, Images, In-Links and so on. Its primary function is to allow highlighting of erroneous links via selection from the Links or Images panels.



Loading Source

Source code is loaded automatically when a page is selected for analysis in the main Results Grid.

SiteScan current supports previewing source code for the following content types:

- HTML, XML and Text
- Style-sheets (e.g. CSS)
- Script files (e.g. JavaScript)

Highlighting Links

To highlight a link in the source code:

1. Select the row for the page to analyze in the main Results Grid.
2. Click the **Links** tab to display all HTML <A> style links.
3. Scroll to the link you wish to highlight and click to select it.
4. All occurrences of the link will be highlighted in the source code using a yellow background.
5. The Source panel caption will display a number showing how many occurrences SiteScan managed to find.

Highlighting Images

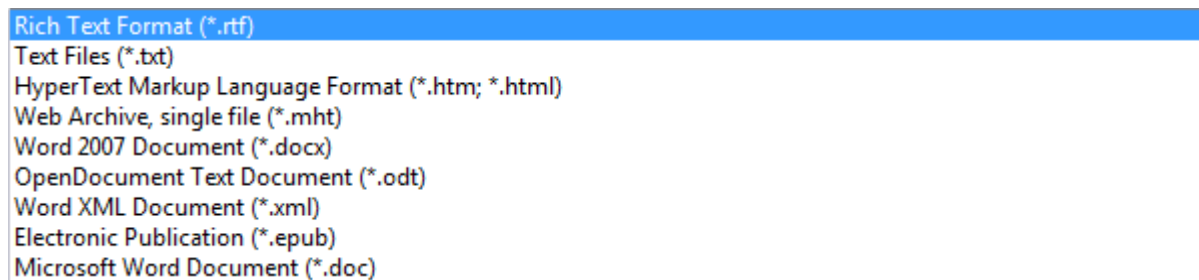
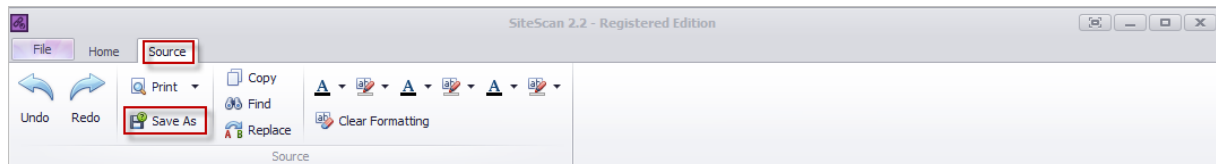
To highlight an image in the source code:

1. Select the row for the page to analyze in the main Results Grid.

2. Click the **Images** tab to display all images on the page.
3. Scroll to the image you wish to highlight and click to select it.
4. All occurrences of the image will be highlighted in the source code using a yellow background.
5. The Source panel caption will display a number showing how many occurrences SiteScan managed to find.

Saving Source

Source code can be saved in various formats by clicking the **Save As** button in the **Source** toolbar ribbon.



Options

Options

[Scanning](#)

[Export](#)

[Licensing](#)

Scanning

Scanning and Performance Options

To display Scanning and Performance options: click the **Options** button on the toolbar ribbon then click the **Scanning** button on the Options pop-up window.

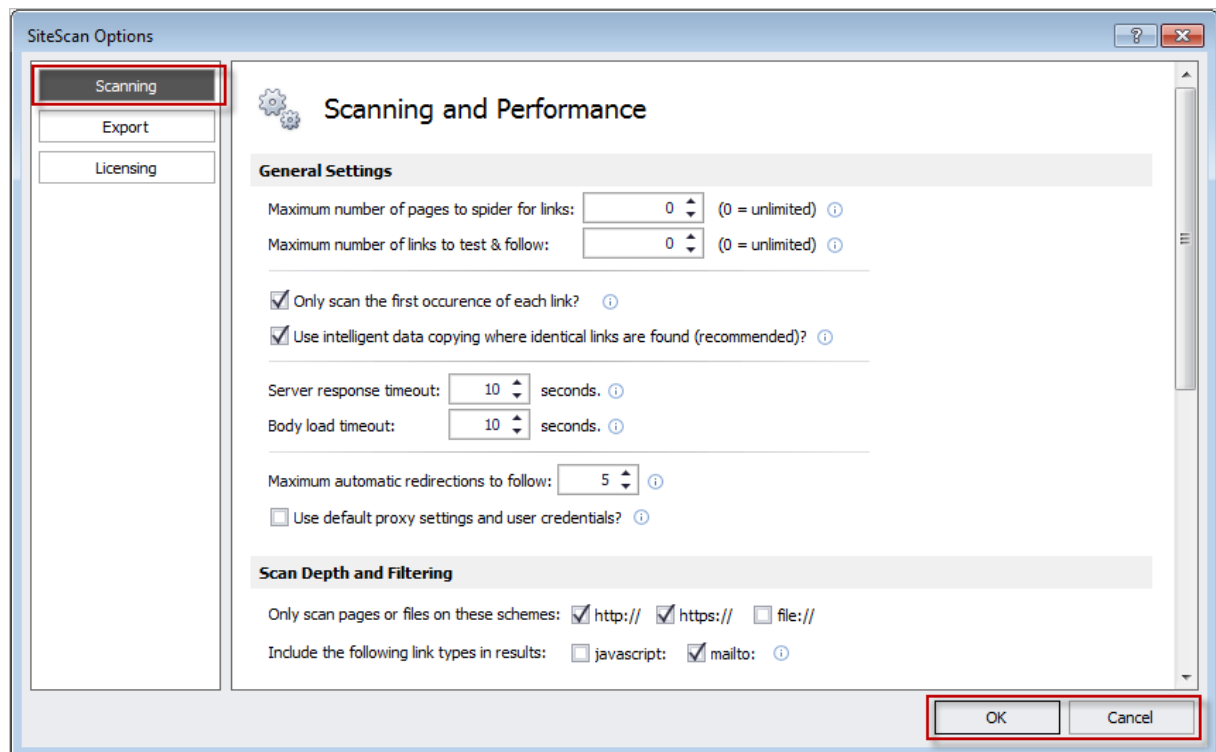
Scanning options should be configured before starting a scan and are saved between SiteScan sessions. Several options can be changed while a scan is in progress and will take immediate effect, such as adding or removing a file extension from an exclusion filter or even changing the scan depth. Options which cannot be changed during a scan are typically those which require the application to be restarted or are related to performance utilization, and will be grayed out while a scan is in progress.

Overview

Scanning options are at the heart of getting the best out of SiteScan; we recommend you explore and experiment to uncover its full potential. Scanning options provide detailed control over the speed and depth of scanning while allowing limits to be set and exclusion filters implemented.

SiteScan ships with all options configured and ready to run: By default it will scan downward from a starting URL and test all links it finds for errors, response time and HTML body load time.

The information icon ⓘ adjacent to each option provides a tool tip summary of the feature along with guidance and best practice, where applicable.



Options in Detail



Scanning and Performance

General Settings

Maximum number of pages to spider for links: (0 = unlimited) ⓘ

Maximum number of links to test & follow: (0 = unlimited) ⓘ

☒ Only scan the first occurrence of each link? ⓘ

☒ Use intelligent data copying where identical links are found (recommended)? ⓘ

Server response timeout: seconds. ⓘ

Body load timeout: seconds. ⓘ

Maximum automatic redirections to follow: ⓘ

☐ Use default proxy settings and user credentials? ⓘ

Scan Depth and Filtering

Only scan pages or files on these schemes: ☒ http:// ☒ https:// ☐ file://

Include the following link types in results: ☐ javascript: ☒ mailto: ⓘ

Only extract and follow links from pages found:

- ☒ On the same level or beneath the Starting URL
- ☐ On the same sub-domain as the Starting URL
- ☐ On the same sub-domain as the Starting URL or any sub-domain

Skip pages or files with the following extensions: ⓘ

Resource Utilization

Scan resource utilization: ⓘ

Give priority to: ☒ Spidering pages ☐ Testing & timing links ⓘ

Timing Measurements

☒ Measure server response time? ⓘ

☒ Measure html document load time? ⓘ

General Settings

Option	Description	Note
Maximum number of pages to spider for links	The maximum number of pages on the website which will be scanned and spidered for links to follow. When set to zero, all pages on the website will be scanned (provided they meet any Scan Depth and Filtering criteria specified)	
Maximum number of links to test & follow	Maximum number of links (found on spidered pages) which will in turn be scanned and spidered for new links to follow.	

	<p>When set to zero, all links found on all pages will be scanned (provided they meet any Scan Depth and Filtering criteria specified).</p> <p>This setting is ignored when Refreshing existing results.</p>	
Scan resource utilization	<p>The maximum number of concurrent web requests which SiteScan will make when scanning pages and spidering links.</p> <p>A higher value will result in greater CPU and memory usage as well as increasing demand on the web server.</p>	<p>Use with caution, too high a value will result in excessive server time-outs due to excessive server load.</p> <p>In link heavy sites, application performance may be impaired. The default value of 20 is a good point. Scale this number up or down to achieve optimum performance with overloading the network or web server.</p>
Give priority to	<p><u>Spidering Pages:</u> Priority is given to link extraction. Select this option when gauging the size of a site or the potential workload is a priority.</p> <p><u>Testing & timing:</u> Priority is given to testing found links by capping the backlog of links waiting to be tested at 500. Select this option when knowing the number of links waiting to be tested is less important than the actual testing.</p>	
Only scan the first occurrence of each page.	<p>Select this option to scan and report on every page in a site only once. Subsequent links to the same page from other pages will be ignored.</p> <p>De-select this option to see all occurrences of links to all pages throughout the site. Ideal when you wish to find all links in a site pointing to a particular page.</p>	
Use intelligent data copying	<p>Select this option to allow SiteScan to copy existing link performance and status data rather than scanning identical links multiple times.</p> <p>When selected both scanning time and web server load are significantly reduced.</p>	
Server response timeout	The length of time in seconds to wait for a response from the web server before skipping.	
Body load timeout	The duration in seconds to wait for the HTML content of a web page to be fully received from the web server before skipping.	
Maximum automatic redirections to follow	The maximum number of automatic server redirections that will be followed before skipping the requested page or file.	
Use default proxy settings and user credentials?	<p>Select this option to use:</p> <ol style="list-style-type: none"> Any proxy settings as configured in Internet Options (via Control Panel). The currently logged on users network credentials for website authentication (if requested by the site). 	<p>The recommended setting for this option is checked.</p> <p>When checked, performance can be reduced if proxy settings are missing or incorrect. To configure proxy settings, see Control Panel > Internet Options.</p>

Scan Depth and Filtering

Option	Description	Note
Only scan pages or files on these schemes	Select the scheme (sometimes referred to as protocol) types you wish to allow when spidering pages and following links.	
Only extract and follow links from pages found:	Select one of the three options to set the depth to which SiteScan will scan. The depth increases with each option and will result in longer scans and larger result sets.	
	An explanation an example of each option follows:	
	<p><u>On the same level or beneath the Starting URL</u></p> <p>If the starting URL is www.mysite.com/products then:</p> <p>1. Only links found at that level or below will scanned.</p>	
	<p><u>On the same sub-domain as the Starting URL</u></p>	

	<p>If the starting URL is www.mysite.com/products then:</p> <ol style="list-style-type: none"> 1. All links found at that level or below will be scanned. 2. All links found adjacent to or above that level will also be scanned. <p>For example, the following links would also be scanned: www.mysite.com/default.html www.mysite.com/services</p>	
	<p><u>On the same sub-domain as the Starting URL or any sub-domain</u></p> <p>If the starting URL is www.mysite.com/products then:</p> <ol style="list-style-type: none"> 1. All links found at that level or below will be scanned. 2. All links found adjacent to or above that level will also be scanned. 3. All links found on any sub-domain of mysite.com will be scanned. <p>For example, the following links would also be scanned: www.mysite.com/default.html www.mysite.com/services products.mysite.com/catalogue</p>	
Skip pages or files with the following extensions	<p>Pages or files with extensions listed here will not be scanned to check if they are responsive or inspected for additional links to follow.</p> <p>Enter a comma separated list of file extensions, such as: .aspx, .exe, .zip, .mp4</p>	
Include the following link types in results	<p>Select any non-navigation type links to also display in scan results.</p> <p>Non-navigable links cannot be checked for a server response.</p> <p>They will be listed in the PageURL column to allow further analysis; such as filtering by a specific email address.</p>	

Performance Measures

Option	Description	Note
Measure server response time	Measure the time it takes for the web server to acknowledge the request for a page or file and return a response (e.g. OK or 404 not found)	
Measure HTML document load time	<p>Measure the time it takes to obtain and load only the HTML <body> content of each web page. All external resources such as images, stylesheets and scripts are ignored.</p> <p>Applies to web pages only. All other file types are automatically ignored.</p> <p>This is a good measure of page performance on the server, particularly for dynamic pages such as ASPX, PHP, etc.</p>	

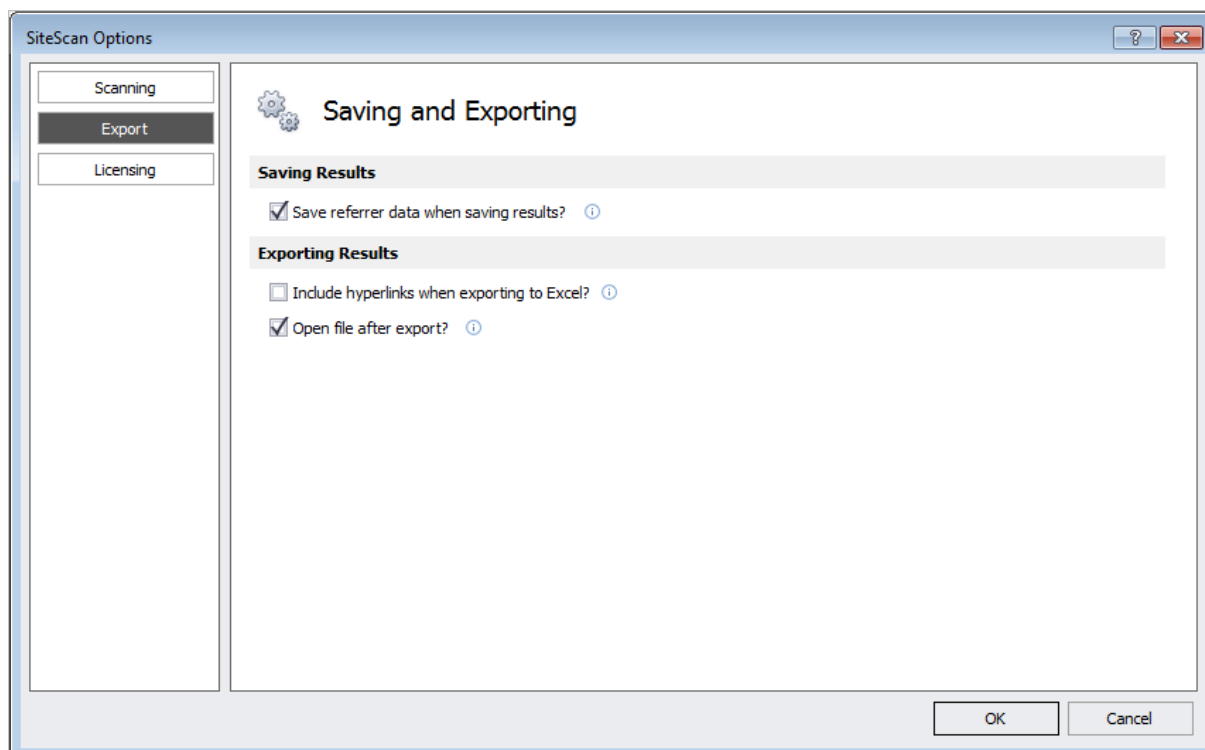
Export

Saving and Exporting Options

To display Saving and Exporting Options: click the **Options** button on the toolbar ribbon then click the **Export** button on the Options pop-up window.

Overview

Saving and Exporting options provide control over how results are saved, exported and allows triggering of convenience actions such as automatically opening results after export.



Option	Description	Note
Save referrer data when saving results	<p>Select this option to also save referrer data whenever scan results are saved.</p> <p>Saving referrer data with results allows viewing of all In-Links found for any page or file when results are later loaded.</p> <p>When selected, loading and saving of results will take longer and produce larger files, particularly on large websites. The referrer data file will be saved alongside the main results file and have a .xmlr file extension.</p> <p>These file should be kept together if sharing results or moving them between computers.</p>	
Include hyperlinks when exporting to Excel	<p>Select this option to create functioning hyperlinks when exporting results to Excel.</p> <p>De-select this option to export hyperlinks as plain text (faster).</p>	<p>Excel can sometimes become unresponsive when attempting to create functioning hyperlinks in result sets which contain restricted content.</p> <p>If you encounter this problem, de-select this option and choose CSV as the export format.</p>
Open file after export	Select this option to launch the newly exported file in Excel (or other native application) when the export has completed.	

Licensing

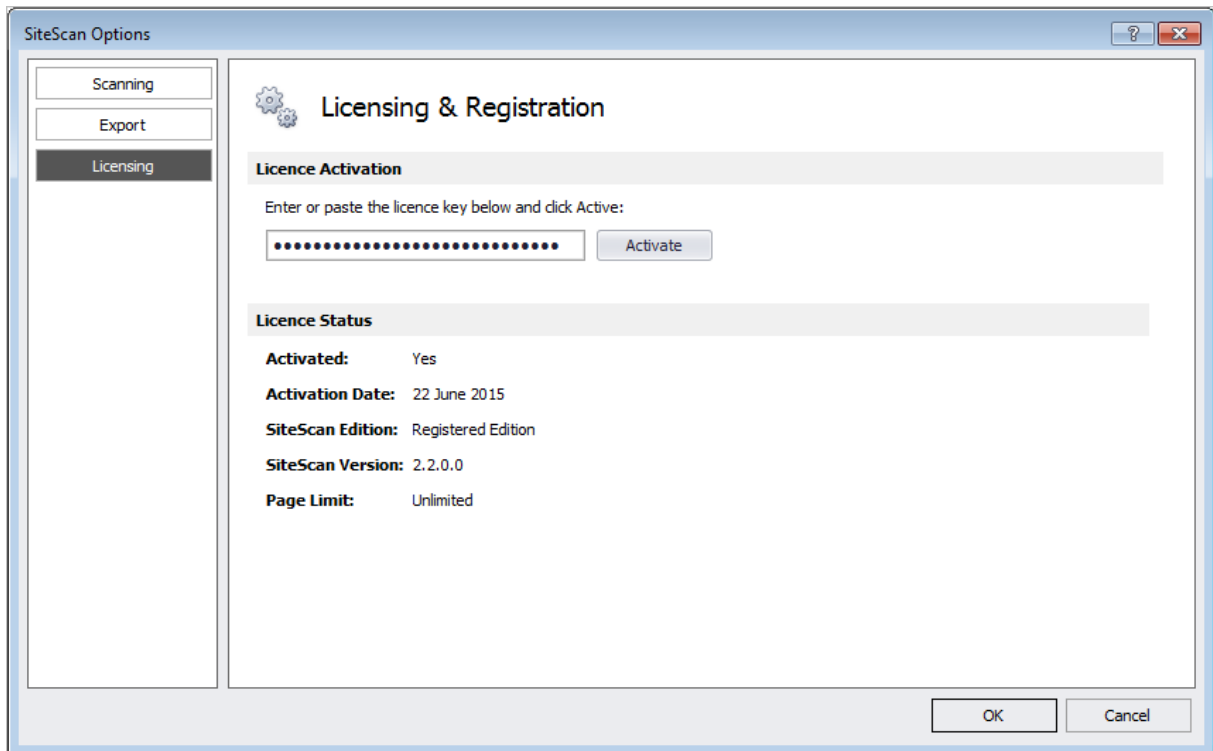
Licensing & Registration

To display the Licensing and Registration screen, either:

- click the **Options** button on the toolbar ribbon then click the **Licensing** button on the Options pop-up window
- click the **Enter License Key** option from the **Buy Now** button on the toolbar ribbon

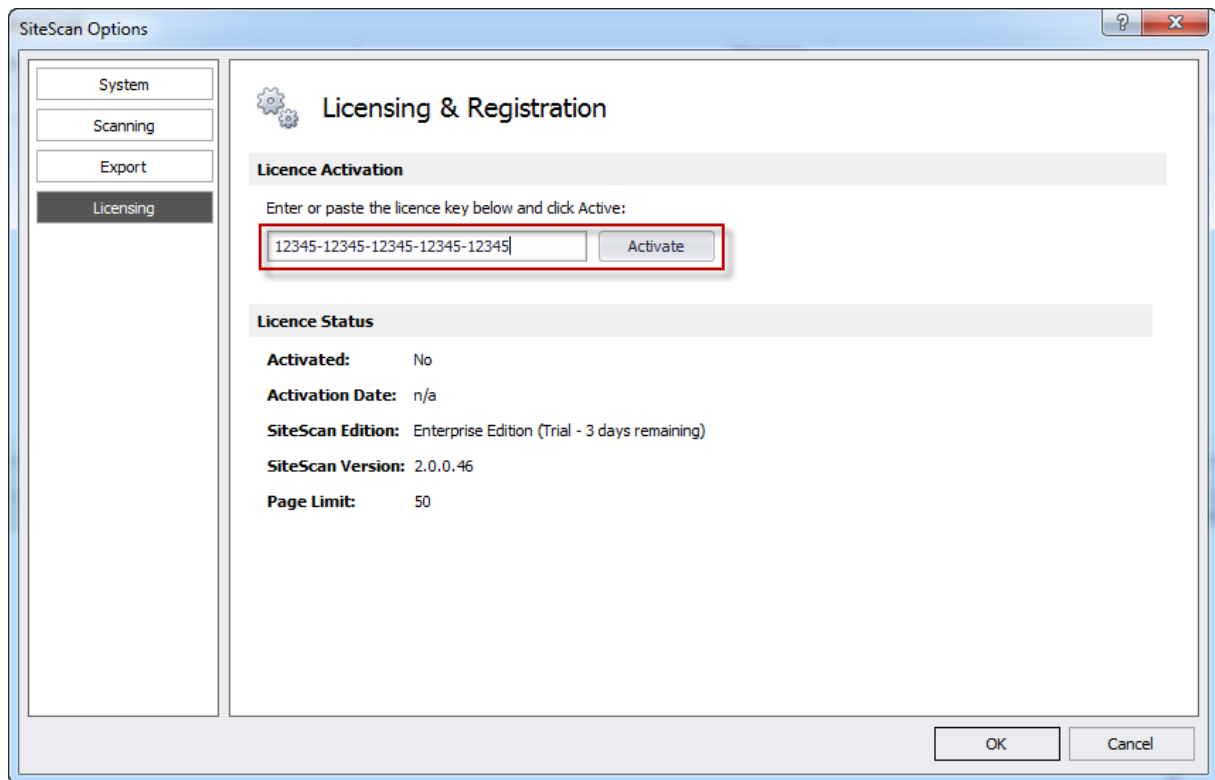
Overview

The Licensing and Registration options screen provides information on the current SiteScan edition and license status. Initial activation and edition upgrades are performed on this screen.



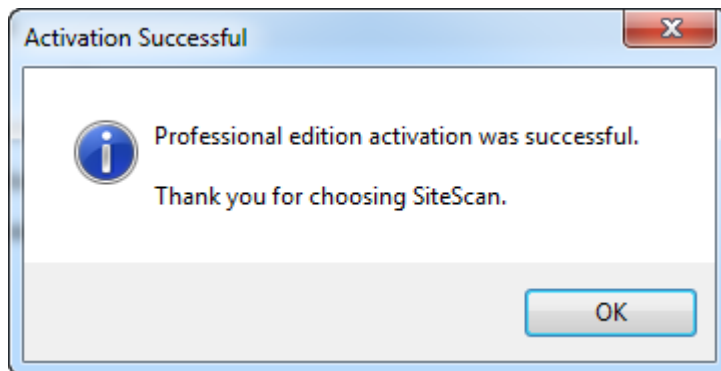
Activating or Upgrading SiteScan

Activation and upgrade follow the same process:



1. Copy the license key from your order confirmation email to the Windows clipboard.
2. Paste the key in to the text box adjacent to the **Activate** button.
3. Click the **Activate** button.

The license key will then be validated and if successful a similar to the following will be displayed:



Current activation status and edition information can be viewed at any time:

